Examining Pediatric Resuscitation Education Using Simulation and Scripted Debriefing

A Multicenter Randomized Trial

Adam Cheng, MD; Elizabeth A. Hunt, MD, MPH, PhD; Aaron Donoghue, MD; Kristen Nelson-McMillan, MD; Akira Nishisaki, MD; Judy LeFlore, PhD; Walter Eppich, MD, MEd; Mike Moyer, MS; Marisa Brett-Fleegler, MD; Monica Kleinman, MD; JoDee Anderson, MD; Mark Adler, MD; Matthew Braga, MD; Matthew Richard, BSc; Vinay M. Nadkarni, MD; for the EXPRESS Investigators

Importance: Resuscitation training programs use simulation and debriefing as an educational modality with limited standardization of debriefing format and content. Our study attempted to address this issue by using a debriefing script to standardize debriefings.

Objective: To determine whether use of a scripted debriefing by novice instructors and/or simulator physical realism affects knowledge and performance in simulated cardiopulmonary arrests.

Design: Prospective, randomized, factorial study design.

Setting: The study was conducted from 2008 to 2011 at 14 Examining Pediatric Resuscitation Education Using Simulation and Scripted Debriefing (EXPRESS) network simulation programs. Interprofessional health care teams participated in 2 simulated cardiopulmonary arrests, before and after debriefing.

Participants: We randomized 97 participants (23 teams) to nonscripted low-realism; 93 participants (22 teams) to scripted low-realism; 103 participants (23 teams) to nonscripted high-realism; and 94 participants (22 teams) to scripted high-realism groups.

Intervention: Participants were randomized to 1 of 4 arms: permutations of scripted vs nonscripted debriefing and high-realism vs low-realism simulators.

Main Outcomes and Measures: Percentage difference (0%-100%) in multiple choice question (MCQ) test (individual scores), Behavioral Assessment Tool (BAT) (team leader performance), and the Clinical Performance Tool (CPT) (team performance) scores postintervention vs pretreatment comparison (PPC).

Results: There was no significant difference at baseline in nonscripted vs scripted groups for MCQ (P = .87), BAT (P = .99), and CPT (P = .95) scores. Scripted debriefing showed greater improvement in knowledge (mean [95% CI] MCQ-PPC, 5.3% [4.1%-6.5%] vs 3.6% [2.3%-4.7%]; P = .04) and team leader behavioral performance (median [interquartile range (IQR)] BAT-PPC, 16% [7.4%-28.5%] vs 8% [0.2%-31.6%]; P = .03). Their improvement in clinical performance during simulated cardiopulmonary arrests was not significantly different (median [IQR] CPT-PPC, 7.9% [4.8%-15.1%] vs 6.7% [2.8%-12.7%]; P = .18). Level of physical realism of the simulator had no independent effect on these outcomes.

Conclusions and Relevance: The use of a standardized script by novice instructors to facilitate team debriefings improves acquisition of knowledge and team leader behavioral performance during subsequent simulated cardiopulmonary arrests. Implementation of debriefing scripts in resuscitation courses may help to improve learning outcomes and standardize delivery of debriefing, particularly for novice instructors.


RESUSCITATION TRAINING PROGRAMS, such as the American Heart Association Pediatric Advanced Life Support (PALS) course, use simulation as an educational modality.1-10 Debriefing following simulated or real resuscitations can improve the process and outcome of resuscitations.10,11 However, the most effective manner in which to train novice instructors to debrief is untested.

CME available online at jamanetworkcme.com and questions on page 502

Currently, PALS instructors complete a certification course, but the quality and
style of instruction remain variable. Few instructors have prior simulation-based education (SBE) or debriefing training.22 Cognitive aids have been used in resuscitation,23 anesthesia,24 and other fields of medicine25,26 to help guide training and care, but to our knowledge, the use of a debriefing cognitive aid for SBE has not been explored. To standardize and improve novice PALS instructor debriefing, we developed a debriefing script as a roadmap for systematic review of trainee performance, focused on existing PALS learning objectives. The script uses language to guide conversation between novice debriefers and trainees and promote reflective learning.

Despite the growing trend of SBE into resuscitation courses, there is little evidence of whether physical realism of simulators affects learning outcomes.11,27,28 The potential addition of high-realism simulation to all American Heart Association courses would be a substantial financial investment for many training centers. The primary objective of this study was to determine whether use of a debriefing script for novice instructors compared with standard debriefing without a script improves PALS-related educational outcomes. The secondary objective was to determine whether use of a high physical-realism simulator “turned on” compared with the same simulator “turned off” (low physical realism) improves PALS-related educational outcomes. Thus, our objective was to determine whether use of a script designed to facilitate debriefings by novice instructors and/or simulator physical realism affects knowledge and team performance of learners in simulated cardiopulmonary arrests.

METHODS

We conducted a multicenter, prospective, randomized, blinded, factorial-design study to assess PALS-related educational outcomes. Research ethics board approval was obtained at all sites. Informed consent was obtained from all participants. Participants were recruited from 14 pediatric tertiary care centers across North America (eMethods). All participants were randomized to 1 of 4 study arms: (1) nonscripted debriefing and low physical-realism simulator, (2) scripted debriefing and low physical-realism simulator, (3) nonscripted debriefing and high physical-realism simulator, and (4) scripted debriefing and high physical-realism simulator.

STUDY PARTICIPANTS

Novice instructors were recruited to debrief simulations. Teams had 4 or 5 participants and were interprofessional. Detailed inclusion and exclusion criteria and team composition are described in the eMethods. All participants (instructors and team members) were distinct and were not recruited to participate in the study multiple times.

STUDY SEQUENCE

Randomization

Each team and novice facilitator was randomized into 1 of 4 study arms (eFigure indicates study flow). All participants were given a standardized orientation to the simulator followed by (1) baseline multiple choice question (MCQ) test, (2) first simulation scenario, (3) debriefing (scripted vs nonscripted) by the novice instructor, (4) second simulation scenario, and (5) post-debriefing MCQ test.

Simulation Scenario

A standardized 12-minute simulation scenario was used that depicted a 12-month-old infant in hypotensive shock progressing to ventricular fibrillation. Two different scenario “stems,” each with different histories of presenting illness (A and B), were written for the same scenario such that participants were unaware that the predebriefing and postdebriefing scenarios were identical (eTable 2).

INTERVENTION

Debriefing Script Development

A debriefing script was designed for novice instructors to facilitate a 20-minute debriefing session (eTables 3-5). The script was developed using an iterative process (eMethods) with a multidisciplinary development team that included pediatric emergency and intensive care physicians, an organizational behavior specialist, a medical educator, and human factors engineers. The language used in the script was based on the debriefing theory known as “advocacy-inquiry.”29,30

Scripted vs Nonscripted Debriefing

All novice instructors received the scenario 2 weeks before the study session. Instructors randomized to scripted debriefing were also given the script but with no instruction on how to use it except with direction to use and follow the script as closely as possible during the debriefing. Instructors randomized to nonscripted debriefing were asked to conduct a debriefing to cover the predefined learning objectives, with no specific instruction on style or method. All instructors held a clipboard while observing the simulation session to access the debriefing script and take notes. This allowed for blinding of the video reviewers as to nonscripted vs scripted debriefing. A research assistant verbally intervened to stop debriefings that extended to 20 minutes.

High vs Low Physical-Realism Simulators

A preprogrammed infant simulator (SimBaby; Laerdal Medical) was used for all simulation sessions. To create a high level of physical realism, full simulator functions were activated (turned on), including vital sign monitoring, audio feedback, breath sounds, chest rise, heart sounds, and palpable pulses.

For editorial comment see page 586

Journal Club slides available at www.jamapeds.com

Low physical-realism groups had the identical simulator but the compressor was turned off, thus eliminating those physical findings. In addition, the low-realism simulator was connected to a monitor, but it displayed only the cardiac rhythm, and not pulse oximetry, respiratory rate, blood pressure, temperature, and audio feedback, which were present in the high-realism group. All other aspects of the simulated resuscitation environment were standardized (eMethods).
OUTCOME MEASURES

Three outcome measures were used: an MCQ test to assess the medical knowledge of individual participants, the Clinical Performance Tool (CPT)\(^3\)\(^1\)\(^,\)\(^2\) to assess the clinical management of the team, and the Behavioral Assessment Tool (BAT)\(^3\)\(^3\)\(^,\)\(^4\) to assess the team leader's behavioral performance. Existing evidence\(^3\)\(^3\)\(^,\)\(^4\) suggests that measures of knowledge, clinical performance, and behavioral performance may be related to changes in patient care and/or outcomes.

Two 20-question MCQ examinations (scored as 0%-100%) were developed using a set of predetermined learning objectives matched to the study scenario and validated as described in the eMethods. The CPT,\(^3\)\(^1\)\(^,\)\(^2\) with 21 individual items (maximum, 42 points; scored as 0%-100%) and designed for evaluation of PALS scenarios, was used to assess clinical performance of the team and validated as described in the eMethods. The BAT is composed of 10 crisis resource management behaviors (maximum, 40 points; scored as 0%-100%). Work by LeFlore et al\(^3\)\(^3\) and LeFlore and Anderson\(^3\)\(^4\) has established the reliability and validity of the tool, with an intraclass correlation coefficient of 0.84 (\(P<.001\)) to 0.95 and a Cronbach \(\alpha\) of 0.95 to 0.97.

RATER TRAINING AND VIDEO REVIEW PROCESS

Sixteen video reviewers comprising pediatric emergency medicine, critical care, and neonatal intensive care physicians and nurse educators rated videos. Each rater was randomly assigned and then blindly viewed and scored 10 to 12 pairs of simulation videos on a password-protected research portal,\(^4\)\(^2\) with each pair representing the presimulation and postsimulation video for a particular team (eMethods).

RANDOMIZATION

Randomization occurred at the level of the team, was stratified by study site, and was conducted in blocks of 4 to ensure equal distribution of teams across the study arms. Randomization packages were prepared at a central study site using a web-based random number generator (http://www.random.org). Sequentially numbered recruitment packages at each site contained 4 opaque envelopes (1 envelope for each study arm) with study arm assignments and random unique identifier codes for the individual participants. One envelope was pulled randomly from the recruitment package for each team on the day of the study. Within each envelope, in addition to study arm allocation, specific assignment of the order of MCQ test delivery (A vs B) for pre- and post-MCQ tests and order of scenario stem delivery (A vs B) for presimulation and postsimulation scenarios were carefully delineated to ensure an even order distribution (A-B vs B-A) among all recruited teams.

STATISTICAL ANALYSIS

All data analysis was performed using statistical software (JMP, version 7.0.1; http://www.jmp.com), with significance designated as \(P<.05\). Pearson \(\chi^2\) was used to assess whether demographics were evenly distributed across study arms. Post-intervention vs preintervention comparison (PPC) scores for MCQ, BAT, and CPT were calculated in percentages. Because each score represents an individual or team compared with itself, this is a form of repeated-measures analysis, with the advantage of limiting intersubject variability. Shapiro-Wilk tests were used to evaluate for normality. The MCQ data were normally distributed, so means with 95% CIs are reported and 1-way analysis of variance was used to test for differences between the 4 study arms. Two-sample independent 1-tailed \(t\) tests (performed on individual PPC scores) queried for differences between scripted vs nonscripted MCQ-PPC and high-realism vs low-realism MCQ-PPC. Because BAT and CPT data were not normally distributed, medians with IQRs were reported, and the Kruskal-Wallis 1-way analysis of variance test was used. Mann-Whitney tests queried for differences between scripted vs nonscripted BAT and CPT-PPC scores and high-realism vs low-realism BAT and CPT-PPC.

RESULTS

STUDY POPULATION

A total of 453 participants composing 104 teams were recruited from July 1, 2008, to February 1, 2011. Of these, 443 individuals (97.8%) completed both the...
pre- and post-MCQ tests and were included in the analysis for that outcome measure. Thirty-seven participants from 8 different teams were randomly selected and removed from our sample to perform further validation of the outcome measurement tools (eMethods). Of the remaining 416 participants (96 teams), 29 participants (6 teams) were dropped from the study because of poor audio or video quality of the recorded simulations (eTable 6). Data from the remaining 387 participants and 90 teams were analyzed for CPT and BAT performance (Figure).

Demographic characteristics of the 443 participants who completed the pre- and post-MCQ tests and 90 novice instructors demonstrated no significant differences between the 4 arms (Table 1 and Table 2).

There was no significant difference in scores between examination A and B (69.6% vs 69.0%; \( P = .80 \)).

**SCRIPTED VS NONSCRIPTED DEBRIEving**

**Knowledge (MCQ Test)**

The mean (95% CI) MCQ test scores did not vary for scripted vs nonscripted debriefing at both baseline (69.3% [67.6-71.1] vs 69.1% [67.4-70.8]; \( P = .87 \)) and postdebriefing (74.6% [73.1-76.3] vs 72.7% [71.1-74.3]; \( P = .09 \)). Participants receiving scripted debriefing showed greater improvement compared with participants randomized to nonscripted debriefing (MCQ-PPC, 5.3% [4.1%-6.5%] vs 3.6% [2.3%-4.7%]; \( P = .04 \)).

### Table 1. Comparison of Demographic Characteristics Between the 4 Study Arms for All 443 Participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Demographic Breakdown</th>
<th>No.</th>
<th>Nonscripted Low Realism (n = 107)</th>
<th>Scripted Low Realism (n = 117)</th>
<th>Nonscripted High Realism (n = 116)</th>
<th>Scripted High Realism (n = 113)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group composition</td>
<td></td>
<td></td>
<td>25 (23.4)</td>
<td>28 (23.9)</td>
<td>26 (22.4)</td>
<td>26 (23.0)</td>
</tr>
<tr>
<td></td>
<td>Team leader</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Team member</td>
<td>338</td>
<td>82 (76.6)</td>
<td>89 (76.1)</td>
<td>90 (77.6)</td>
<td>87 (77.0)</td>
</tr>
<tr>
<td>Occupation</td>
<td>Nurse</td>
<td>204</td>
<td>50 (46.7)</td>
<td>53 (45.3)</td>
<td>55 (47.4)</td>
<td>46 (40.7)</td>
</tr>
<tr>
<td></td>
<td>Paramedic</td>
<td>18</td>
<td>2 (1.8)</td>
<td>4 (3.4)</td>
<td>4 (3.4)</td>
<td>8 (7.1)</td>
</tr>
<tr>
<td></td>
<td>Physician</td>
<td>196</td>
<td>44 (41.1)</td>
<td>51 (43.6)</td>
<td>48 (41.4)</td>
<td>53 (46.9)</td>
</tr>
<tr>
<td></td>
<td>Respiratory therapist</td>
<td>35</td>
<td>11 (10.2)</td>
<td>9 (7.7)</td>
<td>9 (7.8)</td>
<td>6 (5.3)</td>
</tr>
<tr>
<td>MD training level</td>
<td>R1 or R2(^a)</td>
<td>107</td>
<td>23 (25.2)</td>
<td>25 (25.9)</td>
<td>27 (25.4)</td>
<td>30 (25.6)</td>
</tr>
<tr>
<td></td>
<td>R3(^b)</td>
<td>44</td>
<td>13 (29.5)</td>
<td>12 (25.3)</td>
<td>11 (23.4)</td>
<td>8 (15.1)</td>
</tr>
<tr>
<td></td>
<td>R4, R5, or fellow(^c)</td>
<td>44</td>
<td>8 (18.2)</td>
<td>12 (25.3)</td>
<td>9 (19.1)</td>
<td>25 (38.3)</td>
</tr>
<tr>
<td>Non-MD experience</td>
<td>&lt;5 y</td>
<td>154</td>
<td>26 (40.6)</td>
<td>25 (37.9)</td>
<td>20 (29.4)</td>
<td>11 (18.3)</td>
</tr>
<tr>
<td></td>
<td>5-10 y</td>
<td>61</td>
<td>9 (14.1)</td>
<td>14 (21.2)</td>
<td>18 (28.5)</td>
<td>20 (33.3)</td>
</tr>
<tr>
<td></td>
<td>&gt;10 y</td>
<td>115</td>
<td>29 (45.3)</td>
<td>27 (40.9)</td>
<td>30 (44.1)</td>
<td>29 (48.3)</td>
</tr>
<tr>
<td>PALS last taken</td>
<td>&lt;1 y ago</td>
<td>154</td>
<td>30 (28.0)</td>
<td>48 (41.0)</td>
<td>34 (29.3)</td>
<td>42 (37.2)</td>
</tr>
<tr>
<td></td>
<td>1-2 y ago</td>
<td>175</td>
<td>44 (41.1)</td>
<td>35 (29.9)</td>
<td>50 (43.1)</td>
<td>46 (40.7)</td>
</tr>
<tr>
<td></td>
<td>&gt;2 y ago</td>
<td>35</td>
<td>5 (4.7)</td>
<td>14 (12.0)</td>
<td>7 (6.9)</td>
<td>9 (8.0)</td>
</tr>
<tr>
<td></td>
<td>Never</td>
<td>89</td>
<td>28 (20.2)</td>
<td>20 (17.1)</td>
<td>25 (21.6)</td>
<td>16 (14.2)</td>
</tr>
</tbody>
</table>

**Abbreviations:** MD, doctor of medicine; PALS, Pediatric Advanced Life Support.

\(^a\)R1 and R2 indicate postgraduate specialty training year 1 (R1) and year 2 (R2).

\(^b\)R3 indicates postgraduate specialty training year 3.

\(^c\)R4, R5, or fellow indicates postgraduate specialty training year 4 (R4), year 5 (R5), or fellow in subspecialty training (postgraduate year 4 or above).

### Table 2. Comparison of Demographic Composition Between the 4 Study Arms for All 90 Novice Instructors

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Demographic Breakdown</th>
<th>No.</th>
<th>Nonscripted Low Realism (n = 23)</th>
<th>Scripted Low Realism (n = 22)</th>
<th>Nonscripted High Realism (n = 23)</th>
<th>Scripted High Realism (n = 22)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupation</td>
<td>Nurse</td>
<td>19</td>
<td>5 (21.7)</td>
<td>5 (22.7)</td>
<td>4 (17.4)</td>
<td>5 (22.7)</td>
</tr>
<tr>
<td></td>
<td>Respiratory therapist</td>
<td>6</td>
<td>2 (8.7)</td>
<td>1 (4.5)</td>
<td>2 (8.7)</td>
<td>1 (4.5)</td>
</tr>
<tr>
<td></td>
<td>Physician</td>
<td>65</td>
<td>16 (69.6)</td>
<td>16 (72.7)</td>
<td>17 (73.9)</td>
<td>16 (72.7)</td>
</tr>
<tr>
<td>MD training level</td>
<td>R3(^a)</td>
<td>8</td>
<td>2 (12.5)</td>
<td>2 (12.5)</td>
<td>1 (5.9)</td>
<td>3 (18.8)</td>
</tr>
<tr>
<td></td>
<td>R4, R5, or fellow(^b)</td>
<td>57</td>
<td>14 (87.5)</td>
<td>14 (87.5)</td>
<td>16 (94.1)</td>
<td>13 (81.2)</td>
</tr>
<tr>
<td>Non-MD level of experience</td>
<td>&gt;10 y</td>
<td>25</td>
<td>7 (28.0)</td>
<td>6 (24.0)</td>
<td>6 (24.0)</td>
<td>6 (24.0)</td>
</tr>
<tr>
<td>Training program (only for MDs)</td>
<td>Pediatric emergency</td>
<td>18</td>
<td>3 (18.8)</td>
<td>4 (22.0)</td>
<td>5 (29.4)</td>
<td>6 (37.5)</td>
</tr>
<tr>
<td></td>
<td>Pediatric intensive care</td>
<td>16</td>
<td>4 (25.0)</td>
<td>4 (25.0)</td>
<td>4 (23.5)</td>
<td>4 (25.0)</td>
</tr>
<tr>
<td></td>
<td>General pediatrics</td>
<td>26</td>
<td>7 (43.8)</td>
<td>7 (43.8)</td>
<td>7 (41.2)</td>
<td>5 (31.3)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>5</td>
<td>2 (12.5)</td>
<td>1 (6.2)</td>
<td>1 (5.9)</td>
<td>1 (6.3)</td>
</tr>
</tbody>
</table>

**Abbreviations:** MD, doctor of medicine; PALS, Pediatric Advanced Life Support.

\(^a\)R3 indicates postgraduate specialty training year 3.

\(^b\)R4, R5, or fellow indicates postgraduate specialty training year 4 (R4), year 5 (R5), or fellow in subspecialty training (postgraduate year 4 or above).
Team Leader Behavioral Performance (BAT)

Median (interquartile range [IQR]) BAT scores for team leaders did not vary significantly for scripted vs nonscripted debriefing at baseline (52% [38%-71%] vs 54% [40%-67%]; \( P = .99 \)) and postdebriefing (82% [62.5%-90%] vs 74.6% [54.5%-88%]; \( P = .24 \)). Team leaders receiving scripted debriefing showed greater improvement in median BAT scores compared with those receiving nonscripted debriefing (BAT-PPC, 16% [7.4%-28.5%] vs 8% [0.2%-31.6%]; \( P = .03 \)).

Team Clinical Performance (CPT)

Median (IQR) CPT scores did not vary significantly for scripted vs nonscripted debriefing at baseline (73% [68.2%-79.3%] vs 74.6% [69.8%-76.6%]; \( P = .95 \)) and postdebriefing (82.5% [79.3%-87.3%] vs 82.5% [77.7%-85.7%]; \( P = .38 \)). Teams receiving scripted debriefing had improved CPT scores compared with nonscripted debriefing teams (CPT-PPC, 7.9% [4.8%-15.1%] vs 6.7% [2.8%-12.7%]), but this difference was not statistically significant (\( P = .18 \)).

High vs Low Physical-Realism Simulator

There was no significant difference in baseline scores between low-realism and high-realism groups for MCQ (\( P = .24 \)), BAT (\( P = .82 \)), and CPT (\( P = .34 \)). The level of physical realism of the simulator (high realism vs low realism) did not have a statistically significant effect on MCQ-PPC scores (4.9% [3.7%-6.1%] vs 4.0% [2.8%-5.2%]; \( P = .29 \)), BAT-PPC scores (12.0% [6.4%-32.7%] vs 12.7% [0.4%-26.5%]; \( P = .28 \)), or CPT-PPC scores (7.9% [4.8%-14.3%] vs 6.4% [3.2%-12.7%]; \( P = .23 \)). Tables 3, 4, and 5 provide a summary of results for scripting and realism. eTable 7 summarizes secondary

---

**Table 3. Postintervention vs Preintervention Comparison Scores for MCQ, BAT, and CPT**

<table>
<thead>
<tr>
<th>Tool</th>
<th>Postintervention vs Preintervention Comparison Scores</th>
<th>Debriefing</th>
<th>P Value</th>
<th>Realism</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nonscripted</td>
<td></td>
<td>Scripted</td>
<td></td>
</tr>
<tr>
<td>MCQ test (0%-100%),</td>
<td></td>
<td>3.6 (2.3-4.7)</td>
<td>.04</td>
<td>4.0 (2.8-5.2)</td>
<td>.29</td>
</tr>
<tr>
<td>mean (95% CI)</td>
<td></td>
<td>5.3 (4.1-6.5)</td>
<td></td>
<td>4.9 (3.7-6.1)</td>
<td></td>
</tr>
<tr>
<td>BAT (0%-100%), median</td>
<td></td>
<td>8 (0-31)</td>
<td>.03</td>
<td>12 (6-32)</td>
<td>.28</td>
</tr>
<tr>
<td>(IQR)</td>
<td></td>
<td>16 (7-29)</td>
<td></td>
<td>12 (0-28)</td>
<td></td>
</tr>
<tr>
<td>CPT (0%-100%), median</td>
<td></td>
<td>6.6 (3.2-12.7)</td>
<td>.18</td>
<td>7.9 (4.7-14.2)</td>
<td>.23</td>
</tr>
<tr>
<td>(IQR)</td>
<td></td>
<td>7.9 (4.7-14.3)</td>
<td></td>
<td>6.3 (3.1-12.7)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: BAT, Behavioral Assessment Tool; CPT, Clinical Performance Tool; IQR, interquartile range; MCQ, multiple choice question.

**Table 4. Results Before and After Debriefing for MCQ, BAT, and CPT in Scripted vs Nonscripted Debriefing Groups**

<table>
<thead>
<tr>
<th>Tool</th>
<th>Nonscripted Before</th>
<th>Nonscripted After</th>
<th>Scripted Before</th>
<th>Scripted After</th>
<th>P Value Nonscripted vs Scripted After Debriefing</th>
<th>P Value Scripted Before vs After Debriefing</th>
<th>P Value Before vs After Debriefing</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCQ test (0%-100%),</td>
<td>69.1 (67.4-70.8)</td>
<td>72.7 (71.1-74.3)</td>
<td>69.3 (67.6-71.1)</td>
<td>74.6 (73.1-76.3)</td>
<td>.09</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>mean (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAT (0%-100%), median</td>
<td>54 (40-67)</td>
<td>74.6 (54.5-88)</td>
<td>52 (38-71)</td>
<td>82 (62.5-90)</td>
<td>.24</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(IQR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPT (0%-100%), median</td>
<td>74.6 (69.8-76.6)</td>
<td>82.5 (77.7-85.7)</td>
<td>73 (68.2-79.3)</td>
<td>82.5 (79.3-87.3)</td>
<td>.38</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(IQR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: BAT, Behavioral Assessment Tool; CPT, Clinical Performance Tool; IQR, interquartile range; MCQ, multiple choice question.

**Table 5. Results Before and After Debriefing for MCQ Test, BAT, and CPT in High- vs Low-Realism Groups**

<table>
<thead>
<tr>
<th>Tool</th>
<th>Low Realism Before</th>
<th>Low Realism After</th>
<th>High Realism Before</th>
<th>High Realism After</th>
<th>P Value Low vs High Realism After Debriefing</th>
<th>P Value Low Realism Before vs After Debriefing</th>
<th>P Value High Realism Before vs After Debriefing</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCQ test (0%-100%),</td>
<td>68.5 (66.7-70.2)</td>
<td>73.4 (71.8-75.1)</td>
<td>69.9 (68.2-71.6)</td>
<td>73.9 (72.3-75.5)</td>
<td>.67</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>mean (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAT (0%-100%), median</td>
<td>51 (36.5-65.5)</td>
<td>75 (54.5-89)</td>
<td>56 (48-70)</td>
<td>82 (66-88)</td>
<td>.72</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(IQR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPT (0%-100%), median</td>
<td>73 (68.2-76.9)</td>
<td>80.9 (78.9-85.1)</td>
<td>74.6 (69.8-79.3)</td>
<td>84.1 (80.5-87.3)</td>
<td>.10</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(IQR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: BAT, Behavioral Assessment Tool; CPT, Clinical Performance Tool; IQR, interquartile range; MCQ, multiple choice question.
analysis across all 4 study arms. Because realism of the simulation did not have a statistically significant effect on our 3 outcome measures, we did not explore interaction terms (ie, impact of realism + scripted debriefing).

**DISCUSSION**

**SCRIPTED DEBRIEFING**

Our results suggest that novice instructors of a standard PALS course benefit from use of a scripted debriefing tool resulting in improved cognitive and behavioral learning outcomes. Novice instructors, who typically struggle with debriefing aspects of crisis resource management, facilitated discussion of these issues better while using the script, as demonstrated by improved behavioral performance by participant team leaders in simulation following debriefing. Although we did not explore the reason behind this improvement, the positive effect of the script is compelling given the short duration of debriefing and the fact that all instructors were provided learning objectives beforehand.

Our study did not demonstrate a significant improvement in CPT score. The absence of a significant improvement is not unexpected because (1) the CPT is a team-based performance metric, requiring effective interaction of multiple individuals to score positively, and (2) we studied only a single scenario followed by debriefing. The improvement with one scenario and debriefing is encouraging, suggesting that repeated scenario practice and debriefing may improve CPT scores more significantly.

Debriefing is a critically important component of SBE, and many models of debriefing exist. Although cognitive aids have been useful in several medical contexts, we are unaware of any studies assessing the efficacy of a cognitive aid for simulation-based debriefing. Our results support the notion that debriefing is an important element of the simulated learning experience. This work addresses the important issue of instructor competency in standardized resuscitation courses, which rely on large numbers of instructors across many training centers. Recently, the American Heart Association has incorporated a new debriefing tool into the 2011 PALS instructor manuals and courses, signaling a shift in philosophy for instructor training and standardization.

**PHYSICAL REALISM**

Several different ways of categorizing simulation fidelity or realism have been described. For this discussion, we used the categorization of realism into physical, semantical, and phenomenal modes by Dieckmann et al. Physical realism consists of physical properties of the simulator and the environment. Semantical realism refers to concepts, their relationships, and how they influence the simulation. Phenomenal realism includes emotions and cognitive states of thought that people experience during simulation. Several groups have described learning benefits of highly realistic simulation for resuscitation training. Wayne et al demonstrated that inclusion of simulator practice sessions on Advanced Cardiac Life Support algorithms improved cognitive performance in residents as well as adherence to Advanced Cardiac Life Support guidelines in management of actual cardiac arrests. In a randomized trial, Owen et al demonstrated that high-fidelity simulation training compared with low-fidelity training improved cognitive and behavioral performance in medical officers. Lee et al showed that simulator-trained interns performed better trauma assessments compared with interns trained with moulaged patients. In a recent pediatric study, Donoghue et al examined the benefits of physical realism by conducting mock resuscitations for pediatric residents and randomizing them to simulation with physical features activated or inactivated. The residents in the high physical-realism group rated scenarios more highly compared with those in the low physical-realism group, particularly in the pulseless arrest scenarios. Despite growing evidence supporting SBE, to our knowledge no study to date has attempted to describe the specific contribution of physical realism to various types of learning outcomes.

Our study did not demonstrate the same benefits of a high-realism (eg, turned-on) simulator. The effect may have been diminished because other aspects of realism in our study were high. All participants in the study were exposed to a simulated environment with very high physical realism. The preprogramming of scenarios helped to ensure a high degree of semantical realism. In the low-realism groups, phenomenal realism was optimized by using facilitator-guided verbal cues at predefined times. In addition, the scenario we selected for the study involved a pulseless patient, that is, a scenario that did not demand much physical feedback from the mannequin to simulate reality. Thus, our findings related to the lack of simulator physical realism are likely tempered by relatively high degrees of physical, semantical, and phenomenal realism and possibly by selection of a scenario with limited physical findings.

**GENERALIZABILITY**

Although our study was conducted with a specific multidisciplinary group of learners, we suspect that effective use of scripted debriefing would be generalizable to learner groups of varying composition (eg, PALS courses). Scripted debriefing may have a positive effect on more experienced instructors or instructors teaching content in related areas of resuscitation (eg, Advanced Cardiac Life Support). Finally, the impact of a debriefing script may be enhanced with greater familiarity and use of the script. Further research is required to explore the generalizability of scripted debriefing in these related contexts.

**LIMITATIONS**

For practical reasons, we limited the study to one type of scenario, and learners were exposed to only one scenario and one debriefing before the assessment. The brief experience provided by one scenario and debriefing may have introduced a timing bias to our study and limited our ability to assess the full benefit of the intervention. In addition, the debriefing script was provided as a cognitive aid without supplemental instruction on how to
effectively use the tool. This was done to ensure practical application and widespread implementation of the script across American Heart Association training centers. The mode of questioning used in the script is not as open-ended as a traditional reflective debriefing. The phrases in the script were developed to promote some reflection specific to predefined learning objectives but not necessarily to invoke prolonged reflective discussion. Enhanced instructor training and practice using the script before the sessions would likely have altered our results. Furthermore, the debriefing sessions were limited to 20 minutes; thus, the effect of scripted debriefing on variable length debriefings is unknown. Variable adherence to the debriefing script may have affected the results. In some instances, participants managed the simulation unexpectedly, making some of the phrases in the script inapplicable in certain contexts. Because this was difficult to control, we chose to keep all instructors randomized to the scripted debriefing arms in their preassigned arm of the study (ie, intention-to-treat analysis) regardless of how tightly they adhered to the wording of the script.

Our study has demonstrated that scripted debriefing for simulation-based pediatric resuscitation education improves educational outcomes (knowledge) and behavioral performance of the team leader. Turning on or off physical realism features of the mannequin does not improve learning outcomes when other aspects of physical, conceptual, and emotional realism are maintained. Further work is needed to identify the impact of scripted debriefing when used by more experienced instructors, for longer debriefing sessions, and in the context of other types of simulated scenarios.

Accepted for Publication: October 26, 2012. Published Online: April 22, 2013. doi:10.1001/jamapediatrics.2013.1389

Author Affiliations: University of Calgary, KidSim-ASPIRE Research Program, Division of Emergency Medicine, Department of Pediatrics, Alberta Children’s Hospital, Calgary, Alberta, Canada (Dr Cheng); Departments of Anesthesiology and Critical Care Medicine and Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland (Drs Hunt and Nelson-McMillan); Divisions of Emergency Medicine (Dr Donoghue) and Critical Care Medicine (Drs Donoghue, Nishisaki, and Nadkarni), The Children’s Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia; College of Nursing, The University of Texas at Arlington (Dr LeFlore); Division of Emergency Medicine, Ann & Robert H. Lurie Children’s Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, Illinois (Drs Eppich and Adler); TriHealth Education and Simulation Services, Bethesda North Hospital, Cincinnati, Ohio (Mr Moyer); Children’s Hospital of Boston, Harvard Medical School, Boston, Massachusetts (Drs Brett-Fleegler and Kleinman); Division of Neonatology, Doernbecher Children’s Hospital, Oregon Health and Science University (Dr Anderson); Division of Critical Care Medicine, Children’s Hospital at Dartmouth, Hanover, New Hampshire (Dr Braga); Division of Emergency Medicine, Nemours/Alfred I. duPont Hospital for Children, Jefferson Medical College, Wilmington, Delaware (Drs Kost and Stryjewski); Department of Pediatrics, Walter Reed National Military Medical Center, Uniformed Services University of the Health Sciences, Bethesda, Maryland (Drs Min, Podraza, and Lopreiato); Division of Critical Care Medicine, Children’s Hospital of Pittsburgh, Pittsburgh, Pennsylvania (Dr Hamilton); Division of Emergency Medicine, Seattle Children’s Hospital, University of Washington School of Medicine, Seattle (Drs Stone and Reid); Department of Pediatrics, Children’s Medical Center Dallas, Dallas, Texas (Mr Hopkins); Division of Emergency Medicine, Cincinnati Children’s Medical Center, Cincinnati, Ohio (Ms Manos); Division of Critical Care Medicine, Stollery Children’s Hospital, University of Alberta, Edmonton, Alberta, Canada (Dr Duff); and Dementia Guide Inc, Clinical, Halifax, Nova Scotia, Canada (Mr Richard).

Correspondence: Adam Cheng, MD, University of Calgary, KidSim-ASPIRE Research Program, Division of Emergency Medicine, Department of Pediatrics, Alberta Children’s Hospital, 2888 Shaganappi Trail NW, Calgary, AB T3B 6A8, Canada (adam.cheng@albertahealthservices.ca).


EXPRESS Investigators: Kristine Boyle, MS, Lucile Packard Children’s Hospital; John R. Boulet, PhD, Foundation for Advancement of International Medical Education and Research; Laura Corbin, MD, Oregon Health Science University; Marino Festa, MBBS, Children’s Hospital at Westmead; John Gosbee, MD, University of Michigan Health Systems; Laura Gosbee, MD, Red Forest Consulting LLC; Louis P. Halamek, MD, Stanford University;
Takanari Ikeyama, MD, The Children’s Hospital of Philadelphia; Liana Kappus, MD, Yale New Haven Health; Douglas Leonard, MD, Oregon Health and Science University; Frank Overly, MD, Alpert Medical School of Brown University; Jenny Rudolph, PhD, Harvard Medical School; Stephen Schexnayder, MD, University of Arkansas; Robert Simon, EdD, Harvard Medical School; Stephanie Sudikoff, MD, Yale University School of Medicine; and Kathleen Ventre, MD, University of Colorado.

Conflict of Interest Disclosures: Dr Cheng reports receiving a research grant from the American Heart Association (AHA) for design and conduct of the study and collection and analysis of data, a grant from the Laerdal Foundation for Acute Medicine, and an infrastructure grant for the EXPRESS collaborative to support administrative and technical positions (funds from the Laerdal Foundation for Acute Medicine grant were not used for conducting this study). Dr Hunt reports receiving a research grant from the Laerdal Foundation for Acute Medicine, The Hartwell Foundation, and money paid for expert testimony by DeBlasio & Donnell LLC. Dr Nelson-McMillan reports receiving a research grant from the AHA to provide PALS books for this study. Dr Nishisaki reports receiving funding from the Laerdal Foundation Center for Excellence and AHRQ Ro3 Pilot Grant (not relevant to this study). Dr Eppich reports receiving a research grant from the AHA for this study, travel reimbursement as a board member of the Society for Simulation in Healthcare, a research grant from the Agency for Healthcare Research and Quality for an in situ simulation project, and travel reimbursement and salary support from the Center for Medical Simulation and the Tübingen Center for Patient Safety and Simulation. Dr Brett-Fleegler reports receiving a research grant from the AHA. Dr Kleinman reports receiving travel reimbursement as a volunteer for the AHA, is an employee of the Children’s Hospital Boston Anesthesia Foundation, and is a paid speaker for Children’s Mercy Medical Center. Dr Anderson reports being a consultant for SimHealth, employment at Oregon Health Sciences University, receiving payment for a lecture from The Doctor’s Company, and receiving payment from the American Academy of Pediatrics for development of video cases. Dr Kost reports being an employee of Nenomors/Alfred I. duPont Hospital for Children and receiving continuing medical education allocation from employers for travel. Dr Stone reports receiving a research grant from the AHA funding for this study and royalties for Up to Date articles written on femur fractures for children. Dr Reid reports receiving a research grant from the AHA for this study. Dr Nadkarni reports receiving support from the Laerdal Foundation for Acute Care Medicine Center of Excellence Grant (no specific funding for this study), Laerdal Medical Corporation, unrelated research on quality of cardiopulmonary resuscitation and cardiopulmonary resuscitation training, and sponsored travel for visiting professorship and presentation at simulation users groups (unrelated to this study).

Funding/Support: This study was funded by an educational research grant from the AHA.

Role of the Sponsors: Funds from this grant were used for the design and conduct of the study, as well as collection, management, analysis, and interpretation of data.

Online-Only Material: The eTables, eFigure, and eMethods are available at http://www.jamapeds.com. Listen to an author interview about this article, and others, at http://bit.ly/MW1WVH. This article is featured in the JAMA Pediatrics Journal Club. Go to http://www.jamapeds.com to download teaching PowerPoint slides.

Additional Contributions: Anne Marie White, RN, BC Children’s Hospital participated in recruitment and data collection; Albert Ho, Center for Excellence in Simulation Education and Innovation conducted video processing; Ferooz Sekandarpoo, Center for Excellence in Simulation Education and Innovation conducted video processing; and Mary Patterson, MD, Cincinnati Children’s Hospital, participated in study design. No financial compensation was received by these individuals for their contributions to the study.

REFERENCES

17. Shapiro MJ, Morey JC, Small SD, et al. Simulation based teamwork training for emergency department staff: does it improve clinical team performance when
34. LeFlore JL, Anderson M. Alternative educational models for interdisciplinary stu-


